DOCUMENT RESUME

ED 374 158 TM 022 087

AUTHOR Abdel-fattah, Abdel-fattah A.
TITLE Comparing BILOG and LOGIST Estimates for Normal, Truncated Normal, and Beta Ability Distributions.
PUB DATE 94
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ability; *Bayesian Statistics; *Estimation (Mathematics); *Maximum Likelihood Statistics; Monte Carlo Methods; Sample Size; *Statistical Distributions; Test Length
IDENTIFIERS Ability Parameters; *BILOG Computer Program; Data Truncation; Item Parameters; *LOGIST Computer Program

ABSTRACT

The accuracy of estimation procedures in item response theory was studied using Monte Carlo methods and varying sample size, number of subjects, and distribution of ability parameters for: (1) joint maximum likelihood as implemented in the computer program LOGIST; (2) marginal maximum likelihood; and (3) marginal Bayesian procedures as implemented in the computer program BILOG. Normal ability distributions provided more accurate item parameter estimates for the marginal Bayesian estimation procedure, especially when the number of items and the number of examinees were small. The marginal Bayesian estimation procedure was generally more accurate than the others in estimating a, b, and c parameters. When ability distributions were beta, joint maximum likelihood estimates of the c parameters were the most accurate, or as accurate as the corresponding marginal Bayesian estimates depending on sample size and test length. Guidelines are provided for obtaining accurate estimation for real data. The marginal Bayesian procedure is recommended for short tests and small samples when the ability distribution is normal or truncated normal. Joint maximum likelihood is preferred for large samples when guessing is a concern and the ability distribution is truncated normal. Five tables and 27 figures present analysis results. (Contains 30 references.) (Author/SLD)

Comparing BILOG and LOGIST Estimates for Normal, Truncated Normal, and Beta Ability Distributions[1]

Abdel-fattah A. Abdel-fattah

American College Testing, 1994

[1]. The opinions expressed in this article are those of the author and should not be mistaken as representing the policy of American College Testing.

Abstract

The purpose of this study is to compare the accuracy of three estimation procedures in item response theory: the joint maximum likelihood as implemented in the computer program LOGIST; the marginal maximum likelihood; and the marginal Bayesian procedures as implemented in the computer program BILOG. The comparisons were conducted using data generated by a Monte Carlo simulation based on the three-parameter logistic model. The number of items, the number of subjects, and the distribution of ability parameters varied in each simulation. The ability parameter distribution was the variable of most concern.

Normal ability distributions provided more accurate item parameter estimates for the Marginal Bayesian estimation procedure, especially when number of items and number of examinees were small. The Marginal Bayesian estimation procedure was generally more accurate than the other two procedures in estimating a, b, and c parameters. When the ability distribution was beta, the Joint Maximum Likelihood estimates of the c parameters were the most accurate or as accurate as the corresponding Maarginal Bayesian estimates depending on sample size and test length.

Guidelines were provided for obtaining accurate estimation using real data and sample sizes, test lengths, and ability parameter distributions investigated in this study. For example, the Marginal Bayesian procedure is recommended with short tests and small samples for estimating a, b, and c parameters when the ability distribution is normal or truncated normal. The Joint Maximum Likelihood is preferred with large samples when guessing is a concern and the ability distribution is truncated normal.

Comparing BILOG and LOGIST Estimates for Normal, Truncated Normal, and Beta Ability Distributions

Item response theory (IRT) is used in equating, scoring, investigating item bias, and establishing item banks. Consequently, comparing the accuracy of parameter estimates provided by BILOG and LOGIST, the most common IRT programs, is critical to the use of IRT. One feasible way to compare the two programs fairly is to broaden the range of generated values of parameters (Mislevy and Stocking, 1987). This study compares the two programs for three ability distributions using the three-parameter logistic model (3PLM). The 3PLM was chosen for this study because the solution to any problem under this model also applies to the simpler one- and two-parameter logistic models while some solutions for the one-parameter model do not generalize to the two- or three-parameter models. The choice of the more complex model should not undermine the use simpler models because in practice the criterion of choice is model fit not complexity. The 3PLM has the mathematical form

$$P_i(\Theta_j) = c_i + \frac{1 - c_i}{1 + \exp^{f_{ij}}} \qquad (1)$$

where $P_i(\Theta_j)$ = the probability of item i being answered correctly by examinee j, c_i = the lower asymptote of item response curve; $f_{ij} = -1.702a_i(\Theta_j - b_i)$ where $(i = 1,2,3,...,K;\ j = 1,2,3,..., N)$, a_i = discrimination parameter for item i, b_i = difficulty parameter for item i, Θ_j = ability parameter for examinee j, N = number of examinees, and K = number of items.

There are currently four well-founded estimation methods that can be used with all three logistic models: the joint maximum likelihood (JML); the marginal maximum likelihood (MML); the joint Bayesian (JB); and the marginal Bayesian (MB) methods. The JB is only presented for completion and it will not be included in the study due to unavailability of the implementing computer program. The likelihood function for N examinees is the product of N likelihood functions given by the equation

$$L(\Theta;a,b,c) = \prod_{j=1}^{N} \prod_{i=1}^{K} [P_i(\Theta_j)]^{u_{ij}} \, [Q_i(\Theta_j)]^{1-u_{ij}} \qquad (2)$$

where N is the number of examinees, K is the number of items, $P_i(\Theta_j)$ is defined by equation (1) for the 3PLM, and $Q_i(\Theta_j) = 1 - P_i(\Theta_j)$.

Review

Estimation Methods

In this section, the four estimation procedures identified earlier are presented, the literature comparing their accuracy as sample size and test length vary is reviewed, and the literature related to distributional variations in IRT estimation is also reviewed. The JML procedure, the oldest of the four, consists of finding IRT parameter estimates that maximize the likelihood function or its logarithm:

$$\ln [L(\Theta,a,b,c)] = \sum_{j=1}^{N} \sum_{i=1}^{K} [u_{ij} \log P_i(\Theta_j) + (1-u_{ij}) \log Q_i(\Theta_j)] \qquad (3)$$

First the initial values of a_i, b_i, and c_i, are used in estimating Θ_j the unknown parameter. The estimated Θ_j is used in the second stage treating a_i, b_i, and c_i as unknowns to be estimated. This two-stage process is repeated until the ability and item values converge to the final estimates when the difference between estimates of successive stages is negligible. The most commonly known implementation of JML is the program LOGIST developed by Lord (1974). LOGIST has been available since 1973 (Wingersky & Lord, 1973) and has undergone major revision (Wingersky, 1983: Wingersky, Barton & Lord, 1982). The main problem with JML is that item and ability parameters are estimated simultaneously, therefore these estimates may not be consistent. Both item and ability parameter estimates can be consistent for the one-parameter model (Haberman, 1975) and the two- and the three-parameter models (Lord, 1975; Swaminathan & Gifford, 1983) when sample size and test length are large enough. JML ability estimates do not exist for examinees with either perfect or zero scores and JML item parameter estimates for items answered either correctly or incorrectly by all examinees. In LOGIST, the a and c estimates may drift out of bound unless limits are placed on them. For example, Swaminathan and Gifford (1987) placed an upper limit of 2.0 and a lower limit of .06 on the a estimates. However, true JML estimates are not obtained when using restrictions or prior distributions which is the key to Bayesian estimation procedure.

In the joint Bayesian (JB) method (Swaminathan & Gifford, 1982, 1985, & 1986) the likelihood in equation (2) is multiplied by a prior distribution for each of the item and ability parameters to obtain the JB function

$$f(\Theta;a,b,c) = L(\Theta;a,b,c)g(\Theta)g(a)g(b)g(c) \tag{4}$$

The resulting expression is proportional to the joint posterior distribution of these parameters. For example, Swaminathan and Gifford used a normal/gamma/normal/beta prior distribution for the Θ_j, a_i, b_i, and c_i parameters. The use of these or other suitable priors tends to prevent estimates from drifting to intuitively unreasonable values. The authors implemented JB procedure in a computer program that is not currently available for general distribution. A modified JB was implemented in the microcomputer-based program ASCAL (Vale & Gialluca, 1985). The likelihood equations modified for omitted items used in LOGIST were combined with the beta/beta/normal Bayesian prior distributions on a, c, and Θ parameters.

The MML procedure was introduced by Bock and Lieberman (1970). The use of the marginal rather than the likelihood function eliminated the problem of inconsistent item parameter estimates. Multiplying equation (1) by $g(\Theta)$, the probability density function for the ability parameters, and integrating with respect to Θ we obtain the marginal probabilities of the response pattern $\underline{U}$

$$P(\underline{U}) = \int_{-\infty}^{\infty} p(\underline{U} \mid \Theta)\, g(\Theta) \quad d\Theta \tag{5}$$

Once the data are observed this probability can be interpreted as the marginal likelihood function for a given examinee. The product of these likelihoods for all examinees is the marginal likelihood function for the entire data set which can be written as

$$L(a,b,c) = \prod_{j=1}^{N} \int_{\Theta} L(\Theta;a,b,c)g(\Theta) \quad d\Theta \tag{6}$$

The MML estimates are the values of a, b, and c that maximize the likelihood function.

Bock and Lieberman (1970) gave a numerical solution to the likelihood equations. The solution was computationally burdensome and only applicable to tests with 10 or fewer items. Bock and Aitken (1981) refined this solution to avoid computational problems. Mislevy and Bock (1984) implemented this procedure in the program BILOG. In the MML the item parameters are estimated without reference to ability parameters by considering examinees as a random sample from a population and integrating them out of the likelihood function using an approximate ability distribution. For a good approximation of this distribution, a sufficiently large number of examinees is required. Because of this requirement and the integration process, MML involves more computation than does JML. However, MML estimates are more consistent than JML, especially for short tests. Unlike JML, MML has no ability estimates of its own. The maximum likelihood (ML) estimates of abilities can be obtained using MML item parameter estimates and can be abbreviated as ML-MML. The larger the number of items, the better are the ML-MML estimators. As with JML, MML a and c estimates may drift to extreme values. Poor c estimates degrade estimation of other item and ability parameters (Swaminathan & Gifford, 1985). Limits and prior distributions can be used to prevent this drifting. However, these limits and priors produce estimates that are not purely MML. The use of prior distributions introduces the concept of MB estimation.

In the marginal Bayesian (MB) procedure, the likelihood given by equation (6) is multiplied by prior distributions for a, b, and c. The resulting expression is proportional to the posterior density of a, b, and c and can be written as

$$L(a,b,c) = L(a,b,c)g(a)g(b)g(c) \tag{7}$$

MB tends to prevent item parameter estimates from drifting to extreme values. Instead, values are pulled towards the center of the prior distribution for item parameters. That center differs slightly from where it would have been without the priors (Mislevy & Bock, 1984). Therefore, in favorable data, it is preferred to avoid priors entirely and use MML. For unfavorable data, MB of BILOG allows the use of updated or fixed prior means at each iteration. When samples are large relative to the number of items, updated prior means should be used, while for small samples, the fixed prior means are preferred. The default priors in BILOG are lognormal, normal, and beta for a, b, and c; respectively.

Related work on the MB procedure can be found in Dempster, Rubin, and Tsutakawa (1981), Rigdon and Tsutakawa (1983, 1987), and Tsutakawa (1984, 1986). The iterative solution introduced by Dempster et al. (1981) was more general than the similar solution by Bock and Aitken (1981). The latter was limited to random variables with exponential distributions but the former was extended to random variables belonging to non-exponential family distributions. Rigdon and Tsutakawa (1983) derived a marginal maximum likelihood with a fixed b parameter and a random Θ parameter for the one-parameter logistic model by integrating over Θ. This is called the maximum likelihood fixed (MLF) procedure. From the MLF, the conditional maximum likelihood fixed (CMLF) was developed by using the posterior mean of each Θ in the estimation process of the priors to approximate the unknown Bayesian priors conditioned upon their posterior

means. This approximation reduces the computation required by the conventional MML procedure when used in estimating priors. From CMLF, Rigdon and Tsutakawa (1987) derived two more MB procedures under the one-parameter logistic model. These are called the conditional maximum likelihood random (CMLR) and conditional maximum likelihood uniform (CMLU). The prior distribution of b parameter was random in the CMLR procedure and uniform in the CMLU procedure. The ability parameters were assumed random with normal prior distribution. The authors implemented these procedures for the one- and two-parameter logistic models in a computer program unavailable for general distribution.

Sample Size, Test Length, and Estimation Procedure

The JML procedure was found superior to Urry's procedure[2] (e.g., Ree, 1979; Swaminathan & Gifford, 1983). The JML was also found to be superior to the heuristic approximation as implemented in ANCILLES-X (Vale & Gialluca, 1988). Comparing BILOG and LOGIST, Swaminathan and Gifford (1987) concluded that MML for item parameters (or the ML for ability parameters) is generally superior to the JML procedure in estimating a, b, and Θ parameters of the one- and two-parameter logistic models, particularly when small sample size and/or short test lengths were used. For the three-parameter model, LOGIST was superior in estimating b, c, and ability parameters, whereas BILOG was superior in estimating the a parameters. LOGIST estimates of ability parameters were superior because in LOGIST the a parameters were constrained to a reasonable range, the inestimable c parameters were set to a common value, and the program works better with the uniform Θ used in the study. The ML of ability parameters are based on the unconstrained item parameter estimated by MML. The a estimates greater than 4.0 were excluded upon the calculation of the mean squared deviations (MSDs) for both the LOGIST and BILOG estimates; however, these excluded values were greater in number for LOGIST than they were for BILOG.

Using a broader range of generated data , Yen (1987) employed the MB procedure for item parameter estimation and the expected a posteriori (EAP) as well as the ML procedures (ML-MB) for ability estimation using BILOG. She compared these estimates with the corresponding estimates by LOGIST under the three-parameter logistic model. The Θ estimates of EAP were found to be better than either the MB-ML or the JML estimates. Her study was limited to 20- and 40-item tests with 1,000 examinees. Convergence to the true values was investigated only over the increase in test length from 20 to 40 items. In spite of these limitations, BILOG was not superior to LOGIST for all cases. The superiority of LOGIST in some cases might be attributed to the choice of generated values, or to the way the two programs handle extreme estimates. BILOG pulls extreme values towards the center of the prior distribution so that the center differs a little from where it would be if the priors were not used. In LOGIST, upper and lower limits are placed on the a and c parameter estimates to prevent them from drifting to extreme values. Qualls and Ansley (1985) used a limited range of generated data, with various levels of test lengths and sample sizes, under the three-parameter logistic model. They indicated that with ML Θ estimates, the biweight robustification[3] eliminated the problem of assigning the lower-bound ability to high-

---

2. This is the heuristic approximation procedure as implemented in the early version of the computer program ANCILLES.

3. A technique of robust data analysis that improves the accuracy of scale score estimation in the presence of mixed omitting and guessing by scoring omits as incorrect and by giving reduced weight to unlikely correct responses to suppress the effects of guessing.

scoring examinees who missed an easy item. Thus with ability robustification, ability estimation was more accurate with BILOG than with LOGIST.

Swaminathan and Gifford (1982, 1985, & 1986), have shown that the JB estimates are superior to the JML estimates because they do not drift out of range, and are more accurate, even when the prior distributions differ from the distributions of the generated parameters. The JB estimates of ASCAL were also found to be better than the corresponding estimates of LOGIST (Vale & Gialluca, 1985, & 1988). The JB of ASCAL does not provide estimates of the ability parameters, is only available for micro-computers, and takes a long time running large data sets. The JB of Swaminathan and Gifford is not currently available for general distribution.

Consequently it can be concluded that the most important and available procedures for comparisons were the JML of LOGIST and the MB and the MML of BILOG. Precautions were taken so that data generated were reasonable for the two programs. For example, both small and large sample sizes and test lengths were used in the comparison. The JML converges only as both the number of items and the number of examinees increase. The MML and the MB item parameter estimates converge to their true values as the number examinees increases. Thus the small and the large sample-size and test-length combinations were found more important and more reasonable than other combinations. The number of items and the number of examinees chosen in this study were defined as small and large in accordance with some of the aforementioned studies (e.g., Swaminathan & Gifford, 1987).

IRT Parameter Distribution and Estimation Procedures

The JML procedure does not incorporate any assumptions about the distributions of item or ability parameters. The MML procedure requires an assumption about the ability distribution. The JB and MB procedures require assumptions about the priors of both item and ability parameter distributions. There is a small body of literature about the impact of different IRT parameter distributions on the efficiency of various estimations procedures. For example, Swaminathan and Gifford (1983) varied the ability parameter distribution and found it had little effect on JML estimation of the ability and b parameters but did affect estimation of a and c parameters. The a and c estimates were less accurate with negatively skewed ability than with the uniform or normal ability distribution. The uniform ability distribution produced more accurate a and c estimates than the normal ability distribution did. Ree (1979) also found the poorest item parameter estimates with the positively skewed ability parameter distribution and the best item parameter estimates with the uniform ability distribution. The two studies did not include the MML or the MB procedures in the comparison. They both reported differences in accuracy of estimation due to the ability parameter distribution and provided some insight about the importance of varying the ability parameter distribution. The JB procedures were also found to be superior to the JML of LOGIST (Swaminathan & Gifford, 1982, 1985, & 1986) because their estimates do not drift out of range. They were more accurate even when the prior distributions were different from the generated values.

The preceding studies used only the correlations of estimates with the true values except for Yen who used the mean squared deviations (MSDs) as well. The MSD and its component variance and bias provide a means for examining estimates at various levels, while correlations do not. None of the preceding studies provided such comparative measures at several levels of estimates. Swaminathan an Gifford (1987) reported differences of practical interest at several estimate levels but they used only uniform distributions, which worked well with LOGIST, in the three-parameter model. Thus it is important to investigate differences in estimation accuracy across several distributions and to include ability distributions that do not favor one program over another. It is also important to vary the sample size and test length to show convergence across the ability

distributions. The two studies that used large and small numbers of items and numbers of examinees were that by Swaminathan and Gifford (1983) and that of Wingersky and Lord (1985). The two studies did not investigate the BILOG procedures. In the latter study LOGIST estimates were used as the true values.

With the exception of Yen's, none of the preceding studies compared the effect of ability distributions on the estimation accuracy of MB, MML, and JML. Yen (1987) varied the ability parameter distributions and included the JML, MB, ML, and EAP procedures. She held the a and c parameters and the number of examinees constant. The ability parameter distributions used were slightly kurtic and slightly skewed. Therefore, varying the distribution of ability parameters had only a slight effect on the accuracy of the procedures investigated by Yen. The distributions of the b parameters were also varied by Yen (1987) and by Rigdon and Tsutakawa (1987). Rigdon and Tsutakawa recommended the CMLR for small sample size and non-normal b parameter distributions. Because the CMLR program is not available publicly and is restricted to the one- and two-parameter models, the CMLR was not used in this study. The MB procedure of BILOG was used instead.

Among the non-normal ability distributions used in the literature are the uniform and the beta distribution used by Swaminathan and Gifford (1983); the truncated normal distribution used by Ree (1979); and the skewed and the platykurtic distributions used by Yen (1987). The beta and the truncated normal distributions were selected for the present study because these are realistic distributions that showed a negative impact on estimation in previous studies. The uniform distribution is unrealistic and Yen's distribution apparently did not deviate sufficiently from normality to have an effect on estimation accuracy.

Methodology

Design

The conditions varied in this study were the ability parameter distribution (normal, truncated normal, and beta), the test length (20, 60), and the sample size (250, 1000). For each combination of this design the data were replicated 10 times and items of each replication were calibrated by three procedures: JML, MML, and MB. The JML ability estimates of LOGIST were compared to the by-product ML ability estimates from MML (ML-MML) and MB (ML-MB) estimates of BILOG. The total number of data subsets is 120 (2 test lengths x 2 sample sizes x 3 estimation procedures x 10 replications).

Data Generation and Calibration

The data generator used is similar to DATAGEN (Hambleton & Rovinelli, 1973) but is capable of manipulating the IRT parameter distributions as required by this study. The generation process starts with specifying the number of items, the number of examinees, and a suitable seed which produces reasonable ranges of parameter interval. Then, the normal, the beta, and the truncated normal ability parameter distributions are generated; and the beta and truncated normal distributions are standardized[4]. The normal, truncated normal, and beta ability distributions are in

---

4. The mean and standard deviation of the beta distribution were taken from Table II of incomplete beta distributions by Pearson and Hartly (1956, p. 436). The mean and standard deviation of truncated distributions were calculated using the formulae: $\mu = -1.5\ (2\pi)^{-.5}\ (e)^{-.5c}$ and $\sigma = (3/4\ \ [1 + IG(c/2;\ \ \alpha = 1.5,\ \ \beta = 1)] - \mu^2)^{-.5}$, where c is the square of the cut off score (.053), and IG is the integral of the incomplete gamma function with parameters c/2, α, and β. The integral with parameters c/2, α, and β was obtained from Table I of the incomplete Γ-function by Pearson and Hartly (1956, p. 2.).

the ranges (-3.142, 3.020), -1.534, 4.210), and (-3.635, 1.484) respectively. The a_i, b_i, and c_i parameters are generated from lognormal, normal, and beta distributions. The a, b, and c parameters are generated to fall in the ranges (0.363, 2.478), (-2.19, 2.23), and (0.009, 0.343) respectively. Using a_i, b_i, c_i, and standardized Θ_j, the probabilities $P_i(\Theta_j)$ are computed with equation (1). The random numbers X_{ij} are generated from a uniform distribution on the closed interval zero to one. The item responses U_{ij} are generated by comparing X_{ij} with $P_i(\Theta_j)$. If $X_{ij} \leq P_i(\Theta_j)$ then $U_{ij}=1$ otherwise $U_{ij}=0$. Previous steps are repeated to obtain 10 replications for more accurate and stable results.

The generated data are then used as the input for LOGIST and BILOG. The options used in the two programs are the default options. For example, for MB procedure the default priors of BILOG are used. The default number of iteration cycles was fixed in BILOG at 30 for the EM-step and at 6 for the Newton-step. The 60 item test and 1000 examinees (60X1000) was calibrated first. Other subsets (i.e., 60X250, 20X250, and 20X1000) are then calibrated by selecting the specified number of items and number of examinees and running each of the two programs.

Common Metrics for Estimates and True Values

The a, b, and Θ estimates from 120 various BILOG and LOGIST runs were rescaled to be comparable to the corresponding generated true values using the chi-square scaling method described by Divgi (1985). Rescaling is necessary to put the a's, the b's, and the ability estimates of the two programs on the same scale as the corresponding true values (Swaminathan & Gifford, 1987). The equations of linear transformations are $a_{i2}^* = a_{i2}/A$, $b_{i2}^* = A\ b_{i2} + B$, and $\Theta_{j2}^* = A\ \Theta_{j2} + B$, where A is the slope and B is the intercept.

Comparison Indices

The true parameters were compared to the rescaled estimated parameters using the following four criteria of accuracy. These are, the correlation of the estimates with the true parameters, the bias, the variance, and the mean square deviation (MSD) of the estimators. The first order product moment correlation was used to represent relationship between each estimate its corresponding true value. These correlations reflect only linear relationship and they do not reflect the accuracy over replications, therefore the 25th, 50th, and 75th correlation percentiles of the 10 replications were calculated. In addition, the MSD of each estimator from its true value was calculated using the formula

$$\text{MSD} = \sum_{i=1}^{N} (T_i - E_i)^2 \div N \qquad (7)$$

where T_i is the true parameter value for item (or examinee) i, E_i is the estimated parameter for item (or examinee) i, and N is the number of estimated parameters. MSD is the total variance attributed to random and measurement errors. Random or sampling errors relate to the stability of estimation over replications. This component of MSD is called the variance and it can be computed as follows

$$\text{Variance} = \sum_{i=1}^{N} (\bar{E} - E_i)^2 \qquad (8)$$

where $\bar{E}$ is the mean of estimated parameters. The remaining component of the total variance is attributed to errors other than those of sampling fluctuation. This component is called the bias and it can be obtained using the following equation: Bias = MSD - Variance.

Because of the different nature of correlations and MSDs, we will find that conclusions based on correlations sometimes contradicts those based on MSDs. A possible reason for this contradiction is that correlations assume linearity and can be attenuated by nonlinear relationships, but MSDs do not assume linearity. Another reason is that the correlations are reported for only three (median, upper quartile, and lower quartile) out of ten replications, excluding very high or very low correlations while the MSDs is the average deviation for the ten replications including extreme estimates. MSDs get smaller when these estimates are removed or truncated as in JML estimates of the a parameters. For example, JML estimates of a parameters (at least with the 20 items and 250 examinees) had lower correlations than the corresponding MML estimates in spite of the smaller MSDs for the former. Therefore, the best comparison approach is to examine plots together with values of MSDs and to consider correlations more appropriate in the absence of extreme estimates and/or nonlinearity. In other words, for favorable data with no extreme estimates, correlation results are more appropriate because they do not reflect existence of extreme estimates, while for unfavorable data, MSD results are more appropriate. As mentioned earlier, MSD reflect differences at various levels of the estimate scale while correlation does not. Therefore, our discussion and conclusion will focus mainly on the MSD and its components.

Results, Discussion, and Conclusion

The focus of this study is the comparison of LOGIST and BILOG estimates for the normal, truncated normal, and beta ability distributions. Before we examine this comparison, we will briefly compare LOGIST and BILOG estimates within each ability distribution. Within each of the normal and truncated normal ability distribution, MB estimates of a and b parameters were more accurate than the corresponding MML or JML estimates for all sample sizes and test lengths (see corresponding MSDs in Tables 1 and 2). Superiority of MB estimates was more obvious with small sample size and/or short tests. Within the beta ability distribution, the JML estimates of the c parameters were the most accurate or as accurate as the corresponding MB and MML estimates (see MSDs in Table 3). Differences in accuracy between MML and JML depended on the parameter estimated, sample size, and test length. For example, in the 20X250 subset of Table 1 and within each ability distribution, JML had smaller MSDs than those of MML, however bias for MML was smaller and its median correlation was higher than for JML.

Within each ability distribution either ML-MB or ML-MML estimates of ability parameters was more accurate than the corresponding JML estimates for the 20X250 subset. For the other three subsets, ML-MML and ML-MB were more accurate than JML estimates only for some distribution by subset combinations with no obvious pattern (see MSDs Table 4). In addition to differences between estimation procedures within each ability distribution (e.g., MML, MB, and JML for the normal ability distribution), there were differences within each estimation procedure for the three ability distributions (e.g., MML for normal, truncated normal, and beta ability distributions). In the following paragraphs, we will examine differences within estimation procedures for the three ability distributions and we will focus only on large differences in MSDs.

Table 1 shows that the MB estimates of the **a parameters** did not converge to the true values only for one condition when sample size or test length increased. JML estimates did not converge for two conditions while MML estimates did not converge for three conditions. For example, the MSDs of MML estimates increased (from 0.599 to 0.816) for the normal ability distribution when the number of items increased to 60 for the 250 examinees. Nonconvergence for JML estimates occurred with non-normal ability distribution while for MML it occurred with normal ability distribution as well. A possible reason for nonconvergence of MML is the small number of the default iteration cycles in BILOG. The default number of iteration in LOGIST is

larger than in BILOG. Another possible reason is that LOGIST places limits on the a estimate to prevent it from drifting out of range while BILOG does not place any limits on MML estimates.

Within each estimation procedure, MSDs differed for the three ability distributions in three data subsets of Table 1 (notice the underlined MSDs). In the 20X1000 subset, MB estimates had higher MSDs for the beta ability distribution than for other distributions (compare scatterplots in Figure 1). In the 60X250 subset, the MML estimates had lower MSDs for the beta ability distributions than for other distributions (compare scatterplots in Figure 2). In the 60X1000 subset, the MB estimates had lower MSDs for normal distribution than for non-normal distributions (compare scatterplots in Figure 3). While examining the figures, notice that a scatterplot with points scattered away from the agreement (45°) line at part or all of the estimate scale indicate larger MSD than for a scatterplot with points not scattered away from the agreement line.

Table 2 shows that MB estimates of the **b parameters** did not converge to the true values only for two conditions when sample size or test length increased. The JML estimates did not converge for three conditions and MML for five conditions. For example, the MSDs of MML estimates increased (from 0.254 to 0.395) for the normal ability distribution when the number of items increased to 60 for the 250 examinees. Nonconvergence for JML occurred in the non-normal ability distribution while for MML it occurred for normal and truncated normal ability distributions.

Within each estimation procedure, MSDs differed for the three ability distributions in the four data subsets of Table 2 (see the underlined MSDs). In the 20X250 subset, the three procedures had lower MSDs with the beta ability distribution than with other distributions (compare scatterplots in each of the Figures 4-6). In the 20X1000 subset, the MML and MB estimates had lower MSDs with the truncated ability distribution than with other distribution (compare scatterplots in each of the Figures 7 and 8). In the 60X250 subset the results are the same as in the 20X250 subset (compare scatterplots in each of the Figures 9-11). In the 60X1000 subset, MML and MB estimates had lower MSDs with the normal than with the non-normal ability distributions while JML estimates had higher MSDs with the beta than with other ability distributions (compare scatterplots in each of the Figures 12 and 13).

Table 3 shows that the MB estimates for the **c parameters** did not converge to the true values only for three conditions. The MML and the JML estimates did not converge in five conditions each. For example, MSDs increased (from 0.013 to 0.019) for the MML estimates with normal ability distribution when the number of examinees increased to 1000 for the 20-item test.

Within each estimation procedure, MSDs differed for the three ability distributions in the four data subsets of Table 3 (see the underlined MSDs). In the 20X250 subset, MB estimates had higher MSDs with beta than with other ability distributions (compare scatterplots in Figure 14). In the 20X1000 subset, MML estimates had lower MSDs with the normal than with non-normal ability distributions (compare scatterplots in Figure 15) while MB had lower MSDs with beta than with other ability distributions (compare scatterplots in Figure 16). In the 60X250 subset, MB had lower MSDs with normal than with non-normal ability distributions (compare scatterplots in Figure 17) while MML had higher MSDs with truncated normal than with other ability distributions (compare scatterplots in Figure 18). In the 60X1000 subset, MML and MB estimates had higher MSDs for truncated normal than for other ability distributions (compare scatterplots in each of the Figures 19 and 20).

Table 4 shows that the JML estimates of the **ability parameters** converged to the true values. The ML-MML estimates did not converge in four conditions and the ML-MB estimates in six conditions. For example, the MSDs for ML-MML estimates increased (from 0.740 to 1.048) for the normal ability distribution when the number of examinees increased to 1000 for the 20-item test. Possible reasons for this high nonconvergence rate in ML estimates is the limited number of default iterations in BILOG and that the ML estimates are not the best ability estimates in BILOG. Therefore, in practice it is recommended to increase the number of iterations as needed and to use ability estimates other than the ML estimates.

Within each estimation procedure, MSDs differed for the three ability distributions in the four data subsets of Table 4 (see underlined MSDs). In the 20X250 subset, the ML-MB estimates had lower MSDs with the truncated normal than with other ability distributions (compare scatterplots in Figure 21). In the 20X1000 subset, ML-MB estimates had lower MSDs with the normal than with non-normal ability distributions (compare scatterplots in Figure 22). In the 60X250 subset, ML-MML had higher MSDs with truncated normal than with other ability distribution while JML had lower MSDs with the truncated normal than with other ability distribution (compare scatterplots in each of the Figures 23 and 24). In the 60X1000 subset, ML-MML and ML-MB had higher MSDs with the truncated normal than with other ability distribution while JML had lower MSDs with the truncated normal than with other ability distributions (compare scatterplots in each of the Figures 25-27).

Table 5 is a summary of the conditions for which each of the three ability distributions produced more accurate results. As shown in the table, the accuracy of the JML estimation procedure was less affected by varying the ability distribution than other estimation procedures. This effect was expected because JML does not assume the form of the ability distribution while MML assumes normality and MB requires placing prior distributions of a known shape. In agreement with this expectation, it was found that the MML and to a greater extent than MB estimates were more accurate with the normal than with the beta or the truncated normal ability distributions (see the number of conditions for the normal ability distribution). The table also shows that MML and ML-MML worked well with the beta ability distribution while the MB and the ML-MB worked well for a fewer number of conditions with the beta and the truncated normal ability distributions. The MB estimates and the MML estimates converged more often than JML estimates did when the number of examinees (or the number of items) increased. The JML ability estimates converged more often than the ML-MML and the ML-MB did. Convergence was found related to the ability distribution when the a parameters are estimated. Thus estimation accuracy was found to depend on ability distribution, sample size, test length, and estimation procedure.

Accuracy of the item calibration procedure was also found by Ree (1979) to be dependent on the distribution of ability for certain sample sizes and test lengths. Ree's experiment was based on the LOGIST program to calibrate data generated from normal and truncated normal ability distributions. This experiment was extended in this study by including a beta ability distribution and by including MB and ML procedures. In the current study, differences were found not only among estimation procedures but also among the estimates produced with the three ability distributions. These differences were greater with small sample size and/or test length. When both sample size and test length increased, estimates became more accurate and they indicated, in some cases, negligible differences across estimation procedures and across ability distributions. With large samples and long tests, there were appreciable differences between the beta ability distribution and the other two distributions. For example, the MB and the MML estimates of the a and b parameters were more accurate with the normal ability distributions than they were with beta ability distribution. Another example is the ML-MB and the ML-MML estimates which were more

accurate with the beta ability distribution than they were with the other distributions. For many conditions not discussed above differences were negligible among ability distributions.

These negligible differences are in agreement with the results of the study by Yen (1987) who indicated that ability distribution do not affect estimation accuracy. The reason for these negligible differences are in the conditions she investigated as discussed earlier. The appreciable differences can be attributed to the differences in the data generated in the two studies. The results from either of the two studies can be generalized only to data sets with similar conditions. For example, if the a and the c parameters were constant and the ability distribution deviated slightly from normality then it is probable that estimation accuracy will not be affected by ability distribution. On the other hand, if a and c parameters were varied and the deviation from normality was not slight, then the estimation accuracy will be affected by the ability distribution as indicated by Ree (1979) and confirmed in the current study. The less controlled the data were, the more conditions became unfavorable for accurate estimation. Small sample, short tests, varied guessing parameters, and varied discrimination parameters were some of these unfavorable but often uncontrolled conditions. The contribution of this dissertation was to detect differences in accuracy among estimation procedures upon using different ability distributions under these uncontrolled conditions. These differences prevailed even with some favorable conditions. For example, with large samples and long tests, the ML-MML and ML-MB estimates were less accurate at the lower levels for the truncated normal ability distribution (see Figures 25 and 26) than for the normal or the beta ability distributions. These results were based on the MSDs because correlations did not include all replications and they were affected by nonlinear relationships.

The implications of finding differences in accuracy among estimation procedures and among ability distributions are theoretical as well as practical. Theoretically, the effect of varying ability distribution on accuracy of estimation has become evident. Practically, the results about this effect may be used with real data if we know or at least can guess the shape of the true ability distribution. It is true that we do not know the shape of the true ability distribution, however, we may intuitively consider this distribution comparable to the total test scores. For example, after selecting the top two-third from a group of examinees based on their total test cut off score, the true ability distribution can be thought of as truncated normal. Any cut off criterion believed to be correlated with ability can also be used instead of the total test cut off score. Similarly, a normally distributed total test score may be an indicative of a normally distributed true ability. Once we develop a feeling of the shape of the true ability, then we may use the procedure that works best with this distribution for the given sample size and test length as mentioned in Table 5. However, intuition sometimes fail. Therefore, empirical research is needed to help identifying the shape of the true ability distribution probably through a comparison between the distribution type of the total test score or the estimated ability and that of the generated true ability after using an estimation procedure that is least affected by ability distribution.

Table 1. Accuracy Indices for a Parameter Estimates

Estimated Parameter (K X N)	Ability Distribution	Estimation Procedure	Correlation Percentiles 25th	50th	75th	MSD	Squared Bias	Variance
a (20X250)	Normal Θ	MML	.51	.74	.84	0.599	0.104	0.495
		MB	.80	.86	.89	0.178	0.061	0.117
		JML	.33	.53	.56	0.375	0.125	0.250
	Truncated Θ	MML	.45	.67	.77	0.470	0.120	0.350
		MB	.69	.77	.86	0.197	0.077	0.120
		JML	.48	.53	.59	0.350	0.126	0.224
	Beta Θ	MML	.60	.73	.85	0.479	0.094	0.385
		MB	.83	.86	.90	0.317	0.143	0.174
		JML	.39	.54	.57	0.366	0.124	0.242
a(20X1000)	Normal Θ	MML	.89	.89	.90	0.161	0.031	0.130
		MB	.89	.90	.90	0.063	0.021	0.042
		JML	.45	.50	.56	0.371	0.153	0.218
	Truncated Θ	MML	.86	.87	.87	0.156	0.040	0.116
		MB	.86	.88	.88	0.104	0.038	0.066
		JML	.47	.52	.57	0.449	0.201	0.248
	Beta Θ	MML	.86	.87	.88	0.336	0.128	0.208
		MB	.87	.88	.88	0.320	0.154	0.166
		JML	.26	.29	.37	0.508	0.218	0.290
a(60X250)	Normal Θ	MML	.42	.48	.54	0.816	0.193	0.623
		MB	.78	.81	.83	0.125	0.045	0.080
		JML	.57	.64	.69	0.163	0.038	0.201
	Truncated Θ	MML	.39	.44	.53	1.167	0.351	0.816
		MB	.69	.73	.79	0.149	0.050	0.099
		JML	.67	.69	.73	0.150	0.030	0.120
	Beta Θ	MML	.58	.64	.69	0.276	0.059	0.217
		MB	.64	.68	.72	0.213	0.096	0.117
		JML	.55	.62	.66	0.225	0.069	0.156
a(60X1000)	Normal Θ	MML	.86	.89	.90	0.095	0.090	0.086
		MB	.91	.92	.94	0.045	0.015	0.030
		JML	.85	.87	.88	0.091	0.034	0.057
	Truncated Θ	MML	.82	.84	.87	0.174	0.064	0.110
		MB	.82	.82	.89	0.184	0.078	0.106
		JML	.86	.82	.88	0.073	0.026	0.047
	Beta Θ	MML	.81	.84	.88	0.125	0.050	0.075
		MB	.84	.86	.87	0.139	0.064	0.075
		JML	.82	.84	.88	0.122	0.046	0.076

Table 2. Accuracy Indices for b Parameter Estimates

Estimated Parameter (K X N)	Ability Distribution	Estimation Procedure	Correlation Percentile 25th	50th	75th	MSD	Squared Bias	Variance
b(20X250)	Normal Θ	MML	.89	.93	.96	0.254	0.051	0.203
		MB	.93	.97	.98	0.217	0.065	0.152
		JML	.91	.92	.93	1.262	0.575	0.687
	Truncated Θ	MML	.90	.92	.95	0.272	0.095	0.177
		MB	.96	.97	.97	0.262	0.105	0.157
		JML	.89	.91	.92	1.448	0.678	0.770
	Beta Θ	MML	.87	.91	.92	0.484	0.250	0.234
		MB	.93	.95	.95	0.484	0.202	0.282
		JML	.93	.93	.94	2.393	1.120	1.273
b(20X1000)	Normal Θ	MML	.86	.88	.93	0.143	0.031	0.112
		MB	.92	.94	.95	0.120	0.050	0.070
		JML	.97	.97	.98	5.340	2.631	0.709
	Truncated Θ	MML	.85	.86	.86	0.432	0.206	0.226
		MB	.97	.98	.99	0.430	0.205	0.225
		JML	.96	.96	.97	4.338	2.147	2.191
	Beta Θ	MML	.64	.70	.79	0.249	0.076	0.173
		MB	.81	.85	.88	0.231	0.154	0.077
		JML	.97	.97	.97	5.475	2.694	2.781
b(60X250)	Normal Θ	MML	.87	.90	.92	0.395	0.056	0.339
		MB	.96	.96	.97	0.197	0.072	0.125
		JML	.91	.93	.94	0.351	0.093	0.258
	Truncated Θ	MML	.85	.88	.90	0.600	0.188	0.412
		MB	.93	.94	.96	0.354	0.134	0.220
		JML	.90	.92	.94	0.498	0.146	0.352
	Beta Θ	MML	.80	.86	.90	1.180	0.280	0.900
		MB	.95	.96	.97	0.333	0.202	0.282
		JML	.87	.93	.94	1.174	0.383	0.791
b(60X1000)	Normal Θ	MML	.94	.95	.96	0.139	0.023	0.116
		MB	.98	.98	.98	0.093	0.035	0.058
		JML	.97	.97	.98	0.243	0.098	0.145
	Truncated Θ	MML	.92	.94	.96	0.263	0.107	0.156
		MB	.95	.96	.96	0.256	0.118	0.138
		JML	.97	.98	.98	0.246	0.100	0.146
	Beta Θ	MML	.92	.94	.95	0.291	0.095	0.196
		MB	.96	.97	.97	0.205	0.087	0.118
		JML	.96	.97	.97	0.441	0.179	0.262

Table 3. Accuracy Indices for c Parameter Estimates

Estimated Parameter (K X N)	Ability Distribution	Estimation Procedure	Correlation Percentile 25th	50th	75th	MSD	Squared Bias	Variance
c(20X250)	Normal Θ	MML	.41	.50	.58	0.019	0.002	0.017
		MB	.65	.71	.78	0.012	0.004	0.008
		JML	.62	.66	.71	0.013	0.004	0.009
	Truncated Θ	MML	.49	.59	.73	0.020	0.005	0.015
		MB	.67	.72	.84	0.013	0.004	0.009
		JML	.54	.59	.71	0.015	0.004	0.011
	Beta Θ	MML	.45	.48	.54	0.022	0.004	0.018
		MB	.67	.70	.76	0.020	0.009	0.011
		JML	.67	.72	.79	0.012	0.003	0.009
c(20X1000)	Normal Θ	MML	.76	.83	.92	0.010	0.001	0.009
		MB	.81	.88	.94	0.006	0.002	0.004
		JML	.54	.60	.61	0.019	0.008	0.011
	Truncated Θ	MML	.95	.97	.98	0.020	0.008	0.012
		MB	.97	.97	.98	0.010	0.004	0.006
		JML	.47	.60	.70	0.020	0.009	0.011
	Beta Θ	MML	.68	.78	.83	0.020	0.006	0.014
		MB	.73	.80	.85	0.017	0.008	0.009
		JML	.63	.67	.70	0.014	0.006	0.008
c(60X250)	Normal Θ	MML	.45	.52	.63	0.024	0.004	0.020
		MB	.64	.68	.72	0.009	0.004	0.005
		JML	.51	.56	.63	0.015	0.003	0.012
	Truncated Θ	MML	.46	.57	.62	0.036	0.011	0.025
		MB	.61	.64	.68	0.018	0.008	0.010
		JML	.47	.58	.58	0.016	0.003	0.013
	Beta Θ	MML	.33	.42	.48	0.029	0.005	0.024
		MB	.60	.60	.67	0.016	0.007	0.009
		JML	.52	.55	.57	0.016	0.004	0.012
c(60X1000)	Normal Θ	MML	.51	.61	.66	0.015	0.002	0.003
		MB	.74	.76	.78	0.007	0.003	0.004
		JML	.69	.71	.74	0.008	0.003	0.005
	Truncated Θ	MML	.63	.67	.68	0.031	0.014	0.017
		MB	.69	.71	.73	0.020	0.009	0.007
		JML	.68	.69	.75	0.008	0.003	0.005
	Beta Θ	MML	.59	.61	.67	0.019	0.006	0.017
		MB	.73	.77	.80	0.012	0.005	0.011
		JML	.75	.76	.79	0.007	0.002	0.005

Table 4. Accuracy Indices for ability Parameter Estimates

Estimated Parameter (K X N)	Ability Distribution	Estimation Procedure	Correlation Percentile 25th	50th	75th	MSD	Squared Bias	Variance
Θ(20X250)	Normal Θ	ML-MML	.88	.89	.89	0.740	0.232	0.508
		ML-MB	.90	.90	.90	0.748	0.170	0.578
		JML	.75	.76	.77	3.814	0.895	2.919
	Truncated Θ	ML-MML	.83	.84	.85	0.968	0.271	0.697
		ML-MB	.85	.86	.87	1.048	0.238	0.810
		JML	.72	.74	.71	3.800	0.779	3.021
	Beta Θ	ML-MML	.85	.86	.86	0.899	0.301	0.598
		ML-MB	.86	.88	.89	0.867	0.207	0.660
		JML	.77	.79	.79	3.730	1.009	2.721
Θ(20X1000)	Normal Θ	ML-MML	.87	.90	.92	1.048	0.296	0.752
		ML-MB	.88	.91	.94	0.641	0.148	0.493
		JML	.80	.81	.83	1.156	0.244	0.912
	Truncated Θ	ML-MML	.90	.95	.97	1.237	0.305	0.932
		ML-MB	.96	.97	.98	1.221	0.230	0.991
		JML	.77	.80	.80	1.066	0.191	0.875
	Beta Θ	ML-MML	.50	.58	.72	1.061	0.298	0.763
		ML-MB	.76	.81	.84	1.035	0.288	0.747
		JML	.85	.86	.86	1.029	0.253	0.776
Θ(60X250)	Normal Θ	ML-MML	.85	.91	.92	0.541	0.106	0.435
		ML-MB	.94	.95	.95	0.327	0.082	0.245
		JML	.89	.91	.93	0.577	0.131	0.446
	Truncated Θ	ML-MML	.50	.69	.89	1.323	0.282	1.041
		ML-MB	.89	.91	.93	0.468	0.108	0.360
		JML	.94	.96	.96	0.281	0.069	0.212
	Beta Θ	ML-MML	.93	.94	.94	0.339	0.087	0.252
		ML-MB	.94	.95	.95	0.405	0.032	0.373
		JML	.93	.93	.94	0.610	0.199	0.411
Θ(60X1000)	Normal Θ	ML-MML	.94	.94	.95	0.229	0.045	0.184
		ML-MB	.94	.94	.95	0.264	0.058	0.206
		JML	.91	.93	.94	0.422	0.100	0.322
	Truncated Θ	ML-MML	.87	.88	.89	0.570	0.138	0.432
		ML-MB	.87	.88	.88	0.472	0.100	0.376
		JML	.92	.94	.95	0.347	0.083	0.264
	Beta Θ	ML-MML	.94	.94	.95	0.135	0.019	0.116
		ML-MB	.94	.95	.95	0.148	0.026	0.122
		JML	.91	.92	.92	0.618	0.183	0.435

Table 5. Distributions that Produced More Accurate Estimates

Estimated Parameter	(K X N)	Estimation Procedure		
a		MB	MML	JML
	20X250	*	*	*
	20X1000	N,T	*	*
	60X250	N	*	*
	60X1000	*	B	*
b		MB	MML	JML
	20X250	B	B	B
	20X1000	T	T	*
	60X250	B	B	B
	60X1000	N	N	N,T
c		MB	MML	JML
	20X250	N,T	*	*
	20X1000	B	N	*
	60X250	N	T	*
	60X1000	N,B	N,B	*
Θ		ML-MB	ML-MML	JML
	20X250	T	*	*
	20X1000	N	*	*
	60X250	*	N,B	*
	60X1000	N,B	N,B	T

* = negligible effect of ability distribution
B = Beta ability distribution
N = Normal ability distribution
T = Truncated normal ability distribution

BIBLIOGRAPHY

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. Psychometrika, 46, 443-459.

Bock, R.D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored test items. Psychometrika, 35, 179-197.

Dempster, A.P., Rubin, D.B., & Tsutakawa R. K. (1981). Estimation in covariance components models. Journal of the American Statistical Association, 76, 374, 341-353.

Divgi, D.R. (1985). A minimum Chi-square method for developing a common metric in item response theory. Applied Psychological Measurement, 9, 413-415.

Haberman, S. (1975). Maximum likelihood estimates in exponential response models. The Annals of Statistics, 5, 814-841.

Hambleton, R.K., & Rovinelli, R. (1973). A FORTRAN IV program for generating examinee response data from logistic test models. Behavioral Science, 18, 74.

Lord, F.M. (1974). Estimation of latent ability and item parameters when there are omitted responses. Psychometrika, 39, 247-264.

Lord, F.M. (1975). Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters. Princeton, NJ: Educational Testing Service, Research Bulletin 75-33.

Mislevy, R.J., & Bock, R.D. (1984). BILOG Version 2.2.: Item analysis and test scoring with binary logistic models. Mooresville, IN: Scientific Software.

Mislevy, R.J., & Stocking, M. L. (1987). A consumer's guide to LOGIST and BILOG. Princeton, NJ: Educational Testing Service.

Pearson, E.S., & Hartly, H.O. (1956) (2nd ed). Biometrika Tables for Statisticians. London: Cambridge University Press.

Qualls, A.L., & Ansley, T.N. (1985). A comparison of item and ability parameter estimates derived from LOGIST and BILOG. Paper presented at the meeting of the National Council on Measurement in Education, Chicago, IL.

Ree, M.J. (1979). Estimating item characteristic curves. Applied Psychological Measurement, 3, 371-385.

Rigdon, S.E., & Tsutakawa, R.K. (1981). Estimation in latent trait models. (Research Report 81-1). University of Missouri Columbia, Mo 65211. (ERIC REPORTS: ED 208 033).

Rigdon, S.E., & Tsutakawa, R.K. (1983). Parameter estimation in latent trait models. Psychometrika, 48, 4, 567-574.

Rigdon, S.E., & Tsutakawa, R.K. (1987). Estimation of the Rasch model when both ability and difficulty parameters are random. Journal of Educational Statistics, 12, 76-86.

Swaminathan, H., & Gifford, J.A. (1982). Bayesian estimation in the Rasch model. Journal of Educational Statistics, 7, 175-191.

Swaminathan, H., & Gifford, J.A. (1983). Estimation of parameters in the three-parameter latent trait model. In D. Weiss (Ed.), New horizons in testing (pp 14-30). New York, NY: Academic Press.

Swaminathan, H., & Gifford, J.A. (1985). Bayesian estimation in the two-parameter logistic model. Psychometrika, 50, 349-364.

Swaminathan, H., & Gifford, J.A. (1986). Bayesian estimation in the three-parameter logistic model. Psychometrika, 51, 589-601.

Swaminathan, H., & Gifford, J.A. (1987). A comparison of the joint and marginal maximum likelihood procedures for the estimation of parameters in item response models. Draft final report submitted to the Institute for Student Assessment and Evaluation, University of Florida, Gainesville, FL.

Tsutakawa, R.K. (1984). Improved estimation procedures for item response functions. (Final Report on Project NR150-464. Research Report 84-2). University of Missouri Columbia, MO: (ERIC Document Reproduction Service No. ED 250 397).

Tsutakawa, R.K., & Lin, H.Y. (1986). Bayesian estimation of item response curves. Psychometrika, 51, 251-267.

Urry, V.W. (1977). Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 14, 181-196.

Vale, C.D., & Gialluca, K.A. (1985). ASCAL: A microcomputer program for estimating logistic IRT item parameters (Research Report ONR 85-4).

Vale, C.D., & Gialluca, K.A. (1988). Evaluation of the efficiency of item calibration. Applied Psychological Measurement. 12, 53-67.

Wingersky, M.S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R.K. Hambleton (Ed.), Applications of item response theory, (pp 151-156) City: Educational Research Institute of British Columbia.

Wingersky, M.S., Barton, M.A., & Lord F.M. (1982; Version 2.5 updated 1984). LOGIST 5.0 version 1.0 users' guide. Princeton, NJ: Educational Testing Services.

Wingersky, M.S., & Lord, F.M. (1973). A computer program for estimating examinee ability and item characteristic curve parameters when there are omitted responses (RM-73-2). Princeton, NJ: Educational Testing Services.

Wingersky, M.S., & Lord, F.M. (1985). An Investigation of Methods for Reducing Sampling Error in Certain Item Response Theory Procedures. Applied Psychological Measurement, 8, 347-364.

Yen, W.M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. Psychometrika, 52, 275-291.

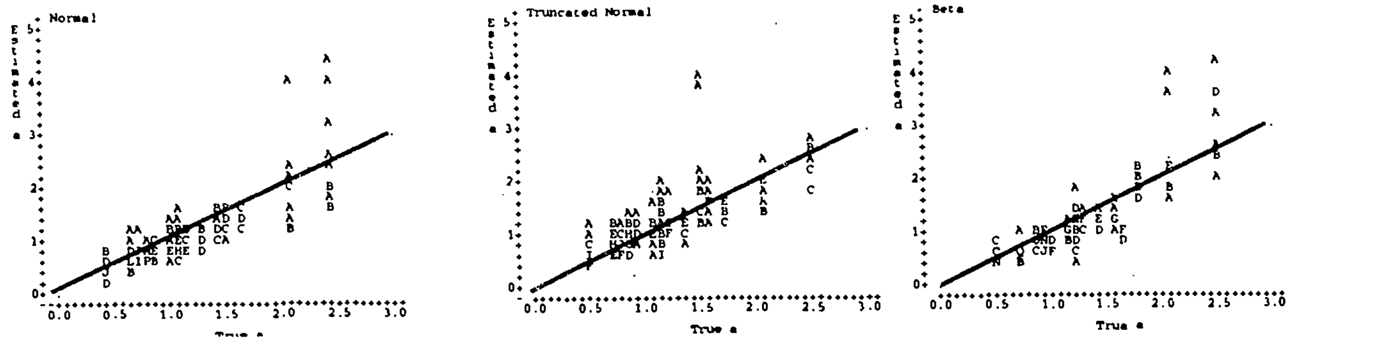

Figure 1. Scatterplots of MML Estimates of a Parameters for 20 Items and 1000 Examinees.

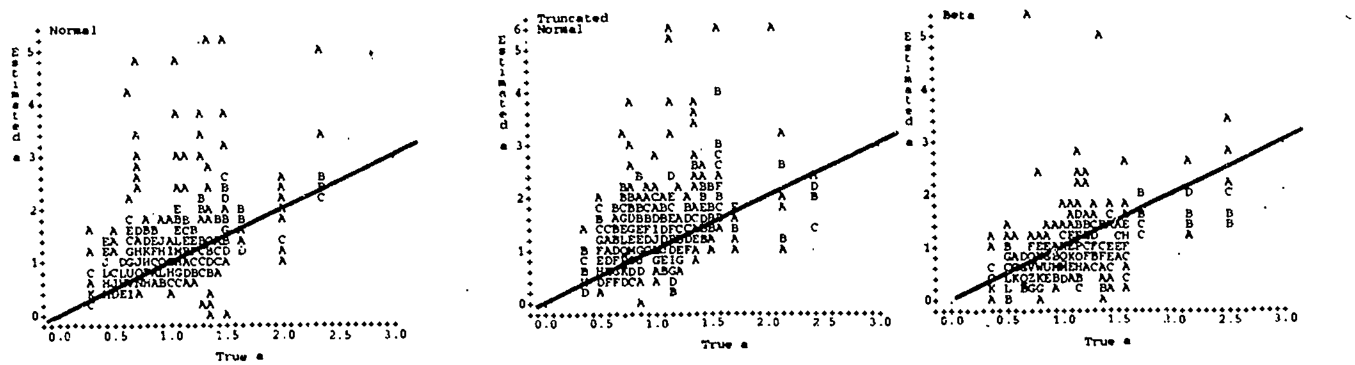

Figure 2. Scatterplots of MML Estimates of a Parameter for 60 Items and 250 Examinees.

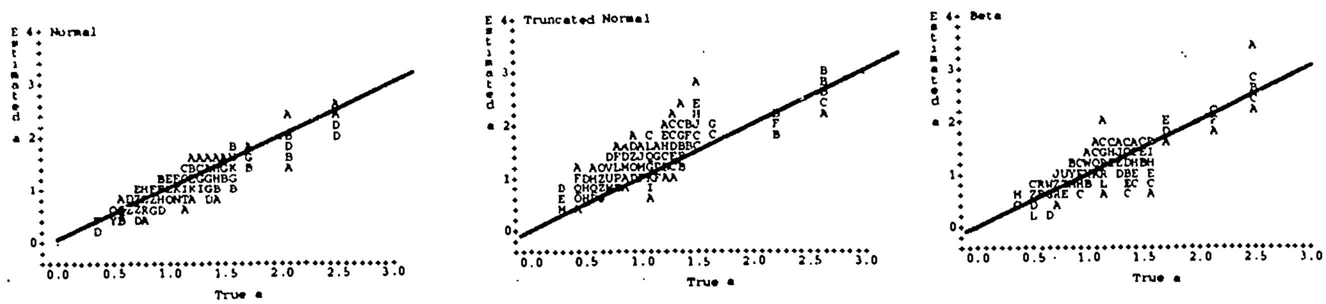

Figure 3. Scatterplots of MB Estimates of a Parameters for 60 Items and 1000 Examinees.

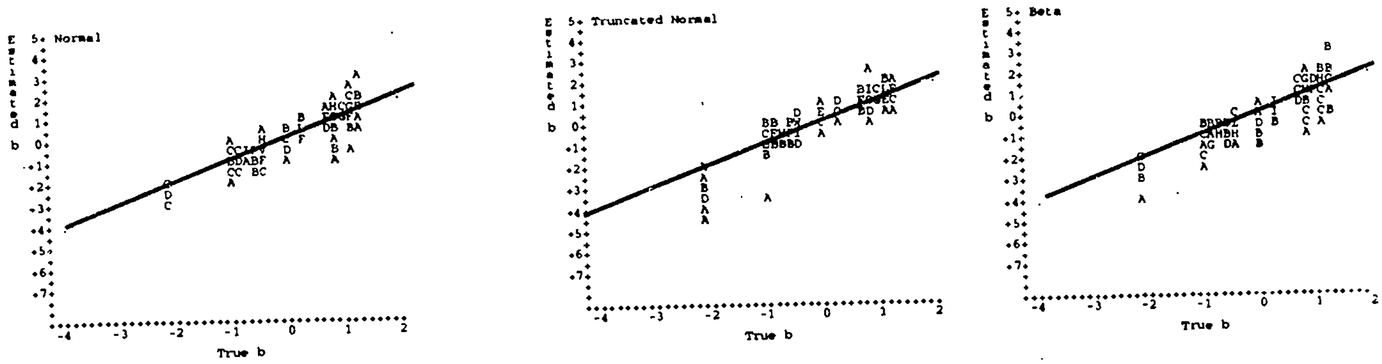

Figure 4. Scatterplots of MML Estimates of b Parameters for 20 Items and 250 Examinees.

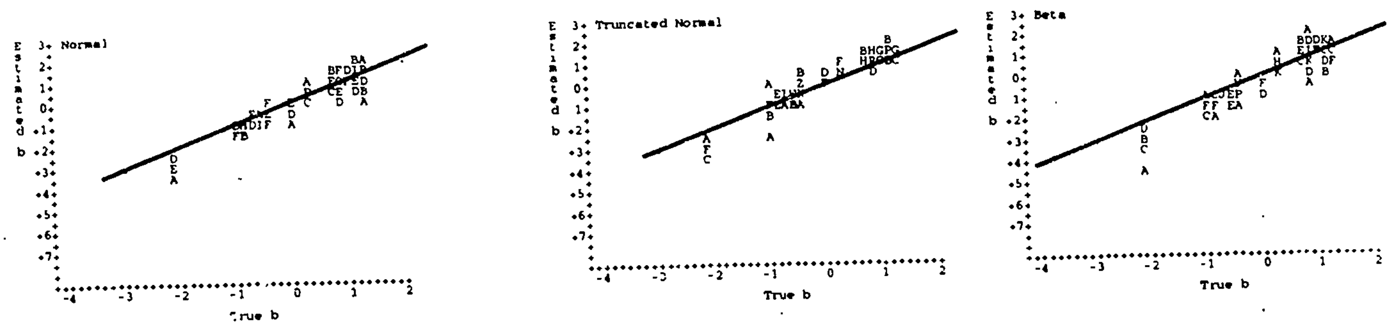

Figure 5. Scatterplots of MB Estimates of b Parameters for 20 Items and 250 Examinees.

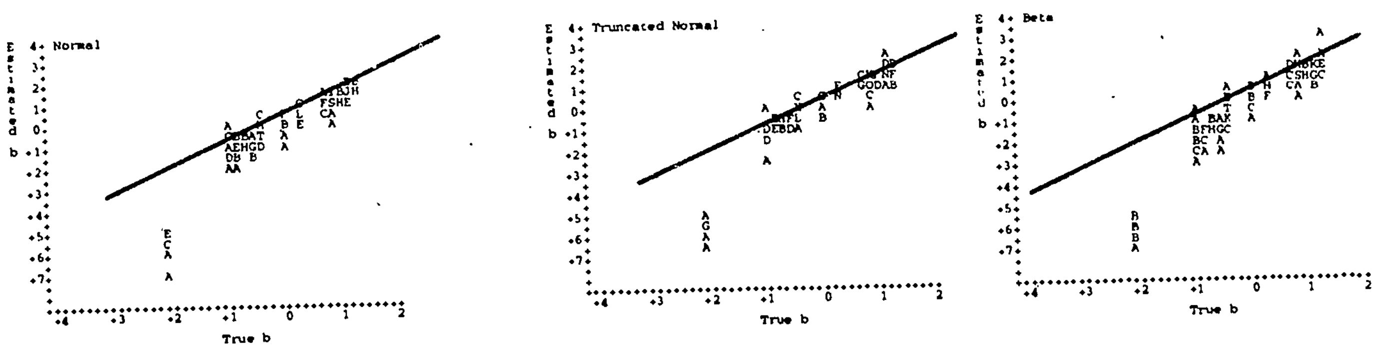

Figure 6. Scatterplots of JML Estimates of b Parameters for 20 Items and 250 Examinees.

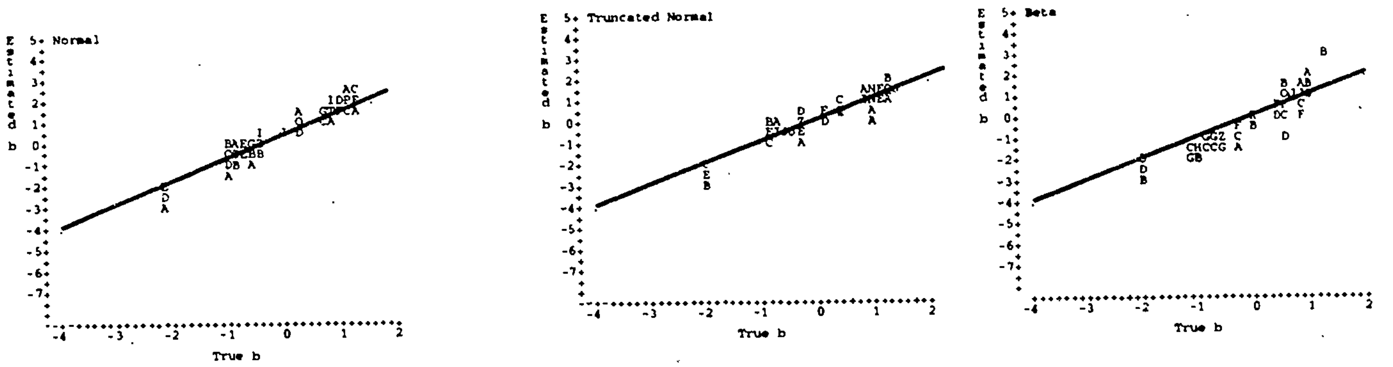

Figure 7. Scatterplots of MML Estimates of b Parameters for 20 Items and 1000 Examinees.

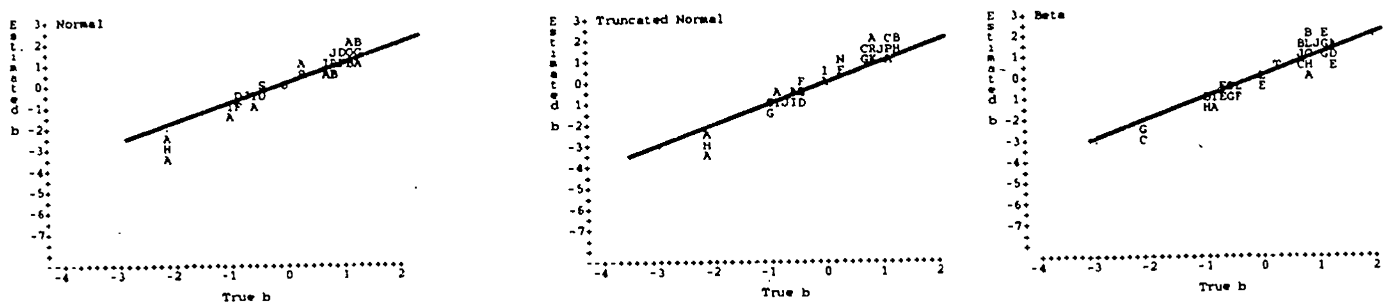

Figure 8. Scatterplots of MB Estimates of b Parameters for 20 Items and 1000 Examinees.

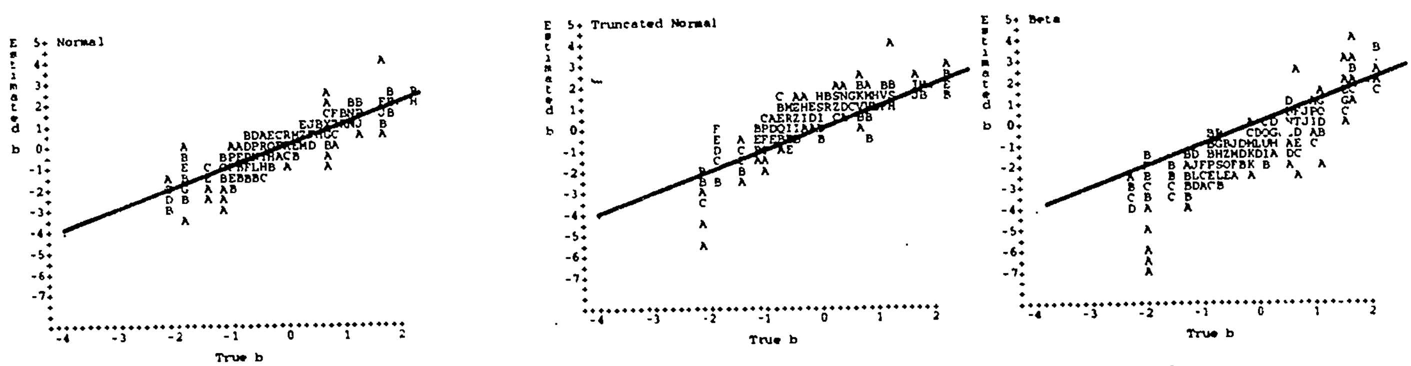

Figure 9. Scatterplots of MML Estimates of b Parameters for 60 Items and 250 Examinees.

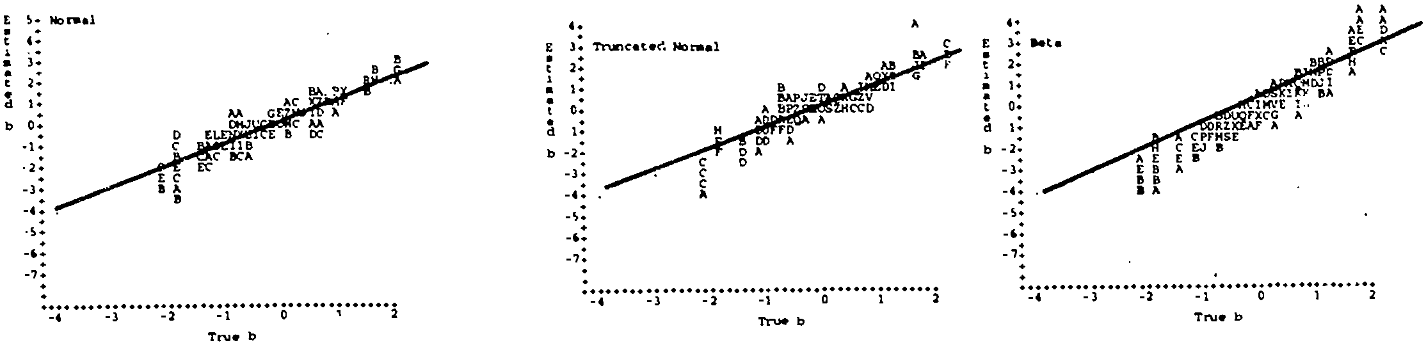

Figure 10. Scatterplots of MB Estimates of b Parameters for 60 Items and 250 Examinees.

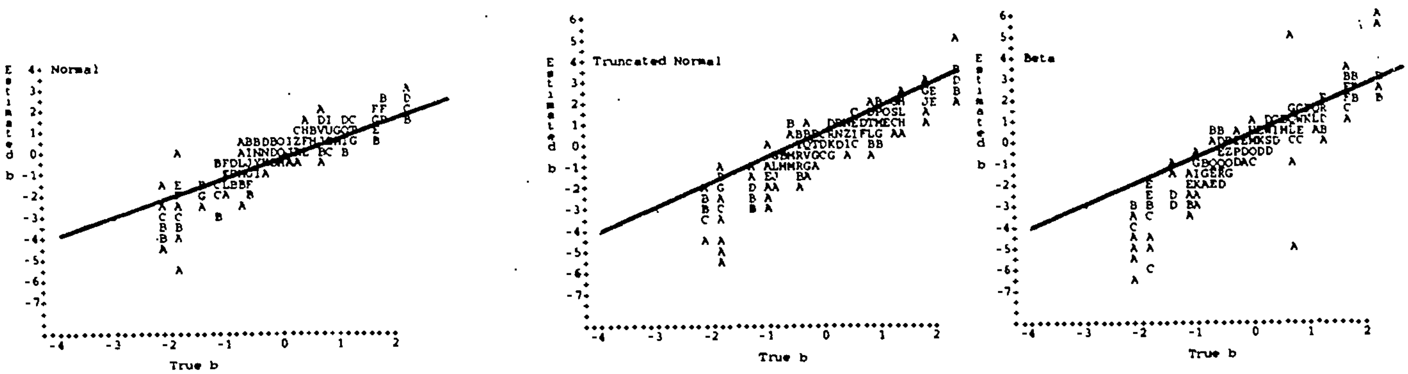

Figure 11. Scatterplots of JML Estimates of b Parameters for 60 Items and 250 Examinees.

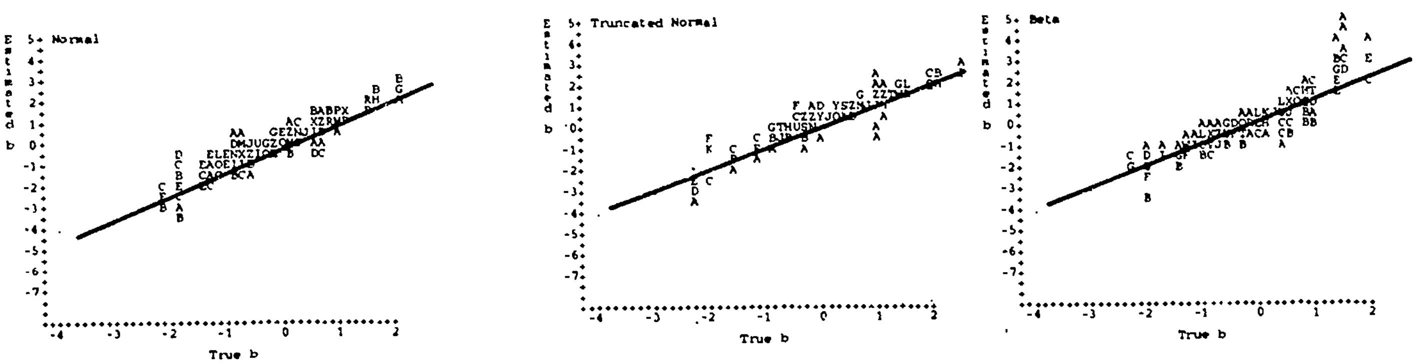

Figure 12. Scatterplots of MML Estimates of b Parameters for 60 Items and 1000 Examinees.

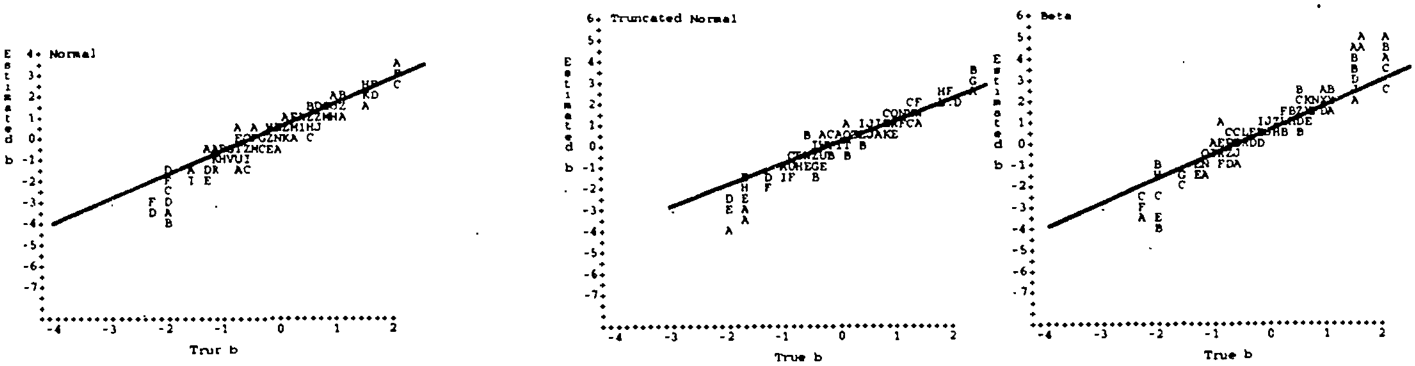

Figure 13. Scatterplots of JML Estimates of b Parameters for 60 Items and 1000 Examinees.

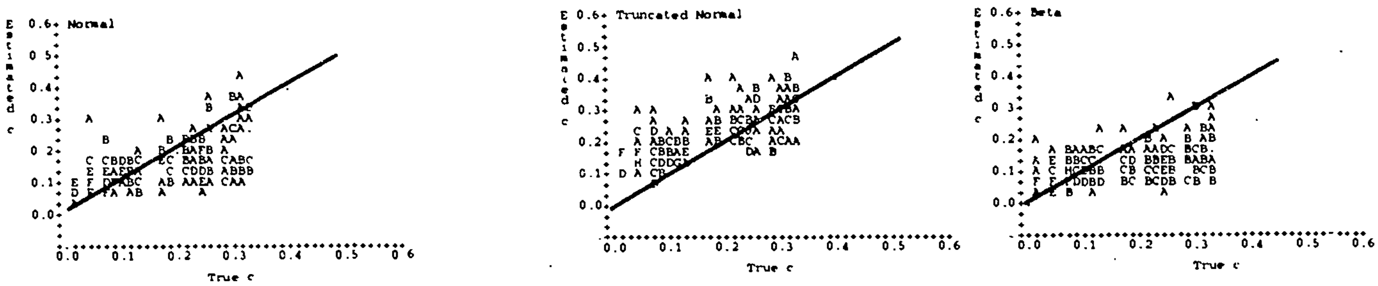

Figure 14. Scatterplots of MB Estimates of c Parameters for 20 Items and 250 Examinees.

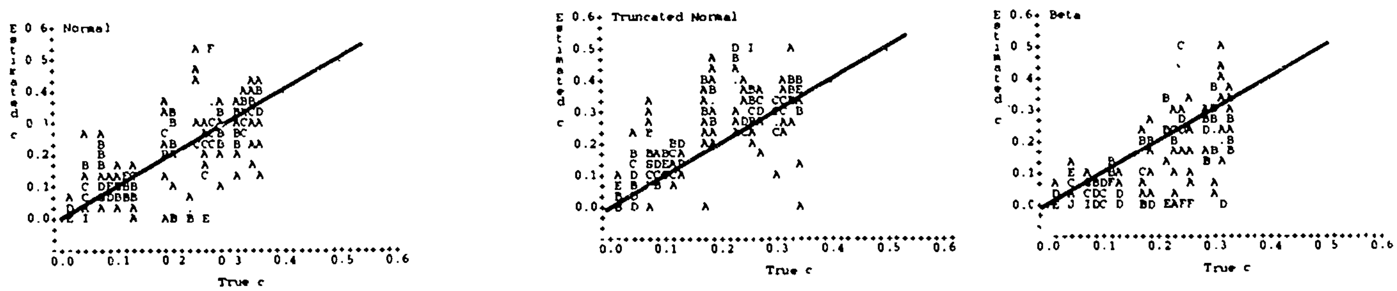

Figure 15. Scatterplots of MML Estimates of c Parameter for 20 Items and 1000 Examinees.

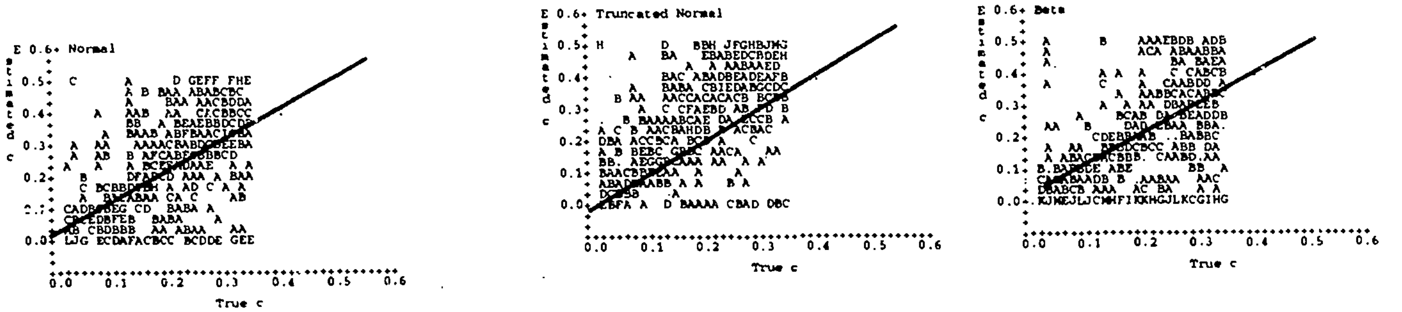

Figure 16. Scatterplots of MB Estimates of c Parameters for 20 Items and 1000 Examinees.

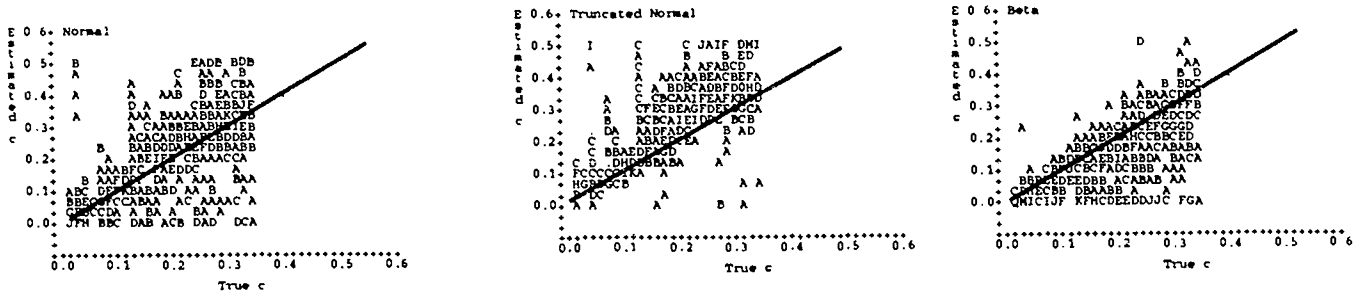

Figure 17. Scatterplots of MML Estimates of c Parameters for 60 Items and 250 Examinees.

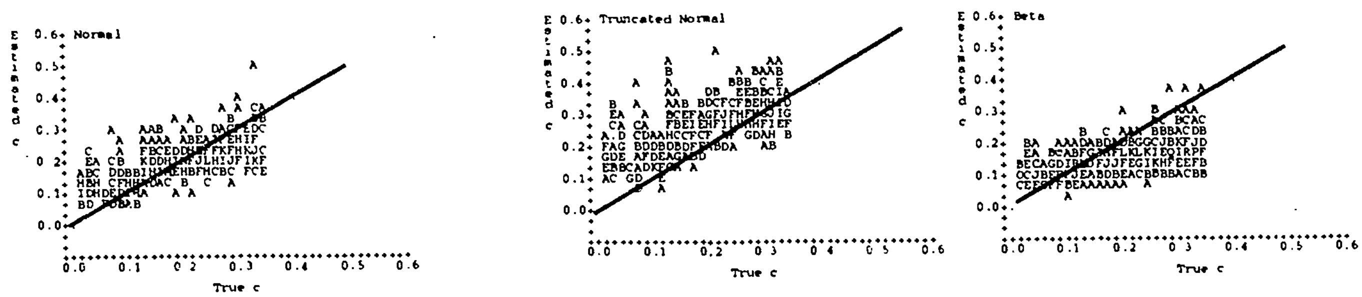

Figure 18. Scatterplots of MML Estimates of c Parameters for 60 Items and 1000 Examinees.

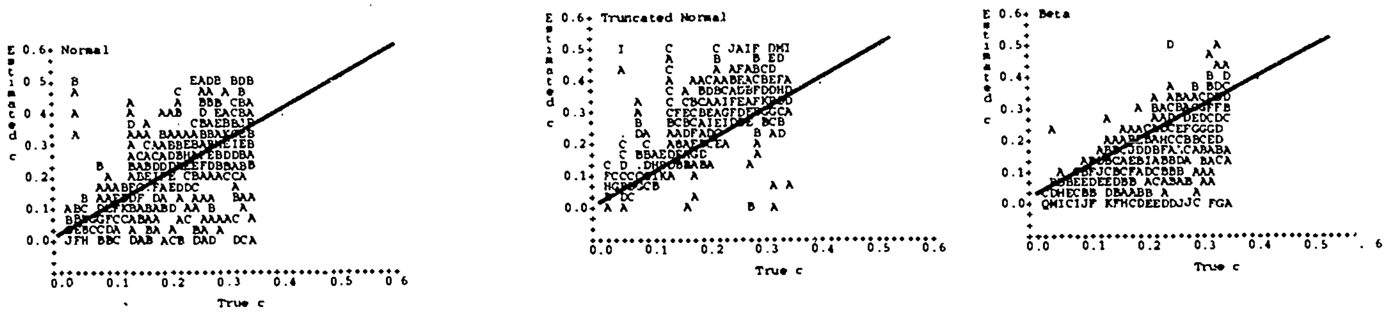

Figure 19. Scatterplots of MML Estimates of c Parameters for 60 Items and 1000 Examinees.

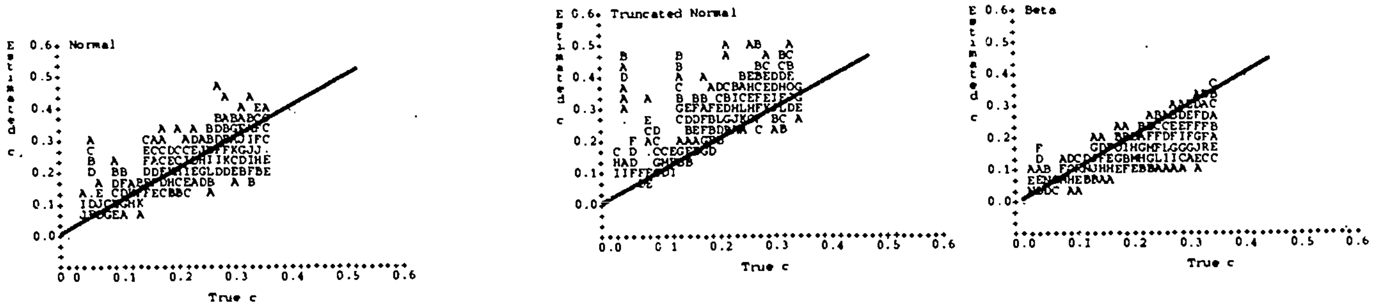

Figure 20. Scatterplots of MB Estimates of c Parameters for 60 Items and 1000 Examinees.

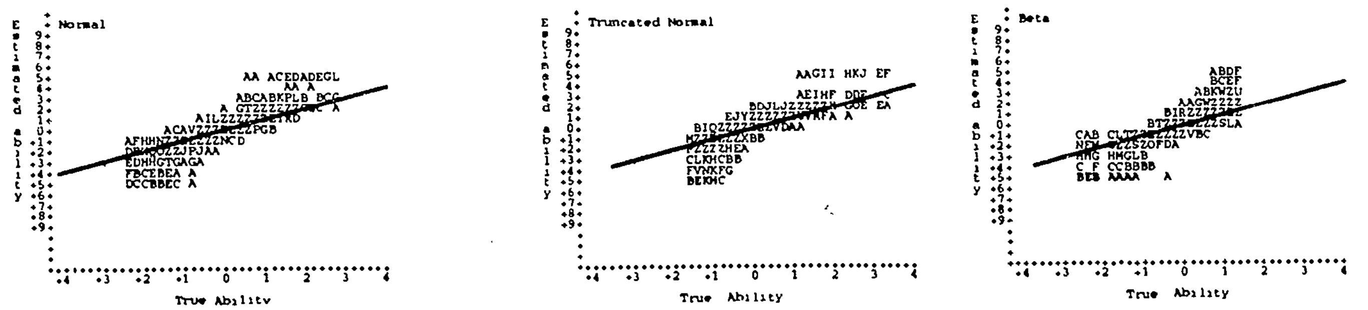

Figure 21. Scatterplots of ML-MB Estimates of Θ for 20 Items and 250 Examinees

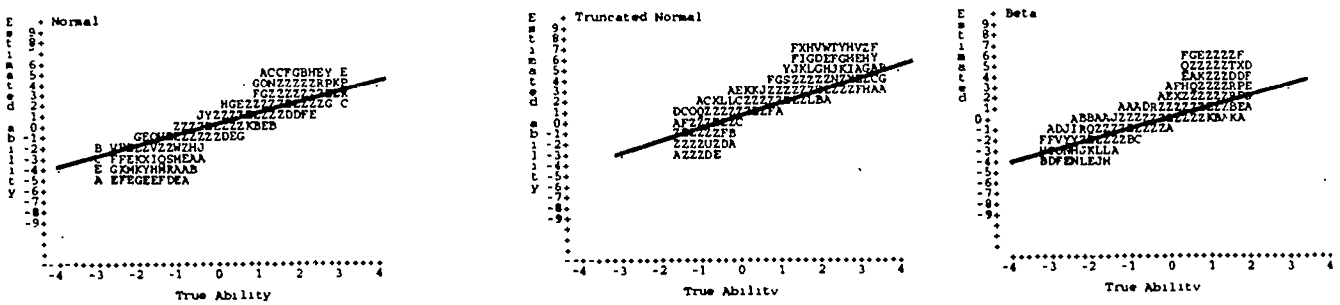

Figure 22. Scatterplots of ML-MB estimates of Θ for 20 items and 1000 examinees.

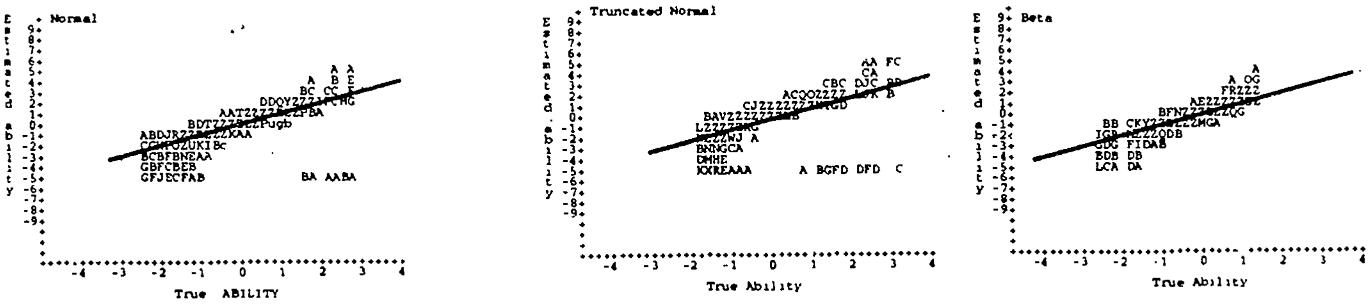

Figure 23. Scatterplots of ML-MML Estimates of Θ for 60 Items and 250 Examinees.

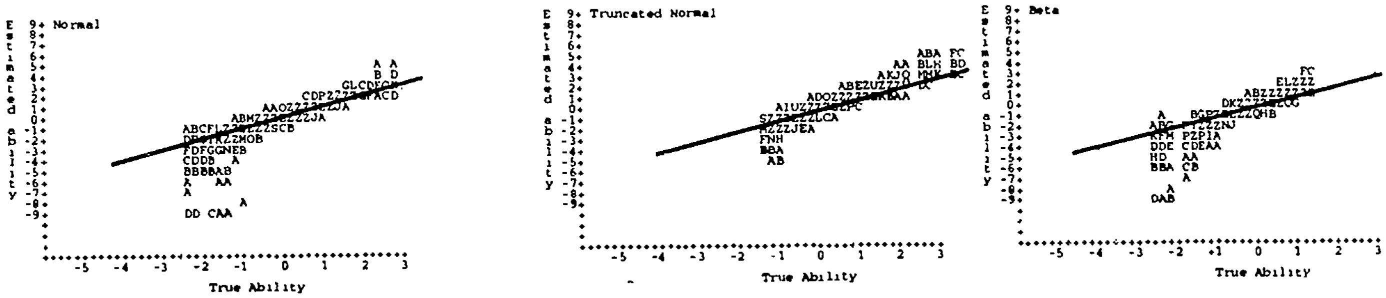

Figure 24. Scatterplots of JML Estimates of Θ for 60 Items and 250 Examinees.

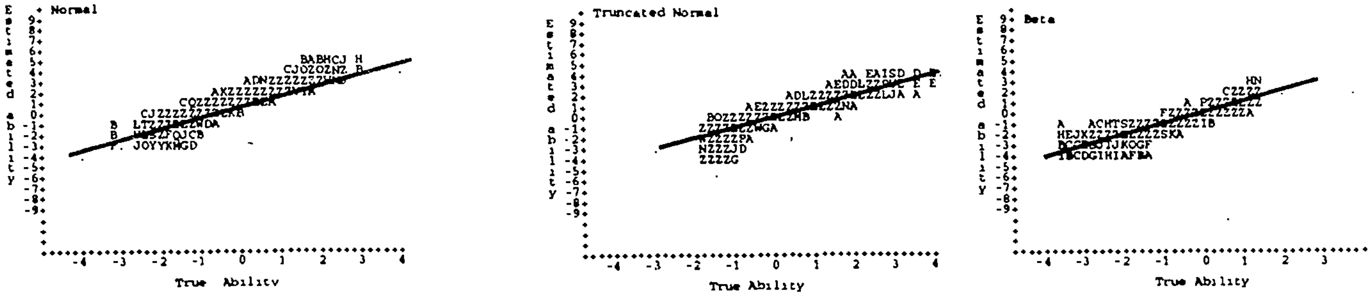

Figure 25. Scatterplots of ML-MML Estimates of Θ for 60 items and 1000 examinees.

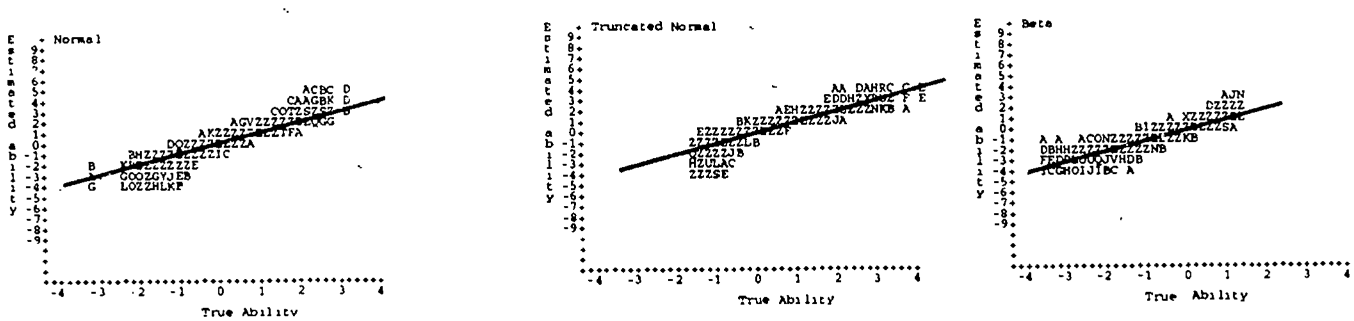

Figure 26. Scatterplots of ML-MB estimates of Θ for 60 items and 1000 examinees.

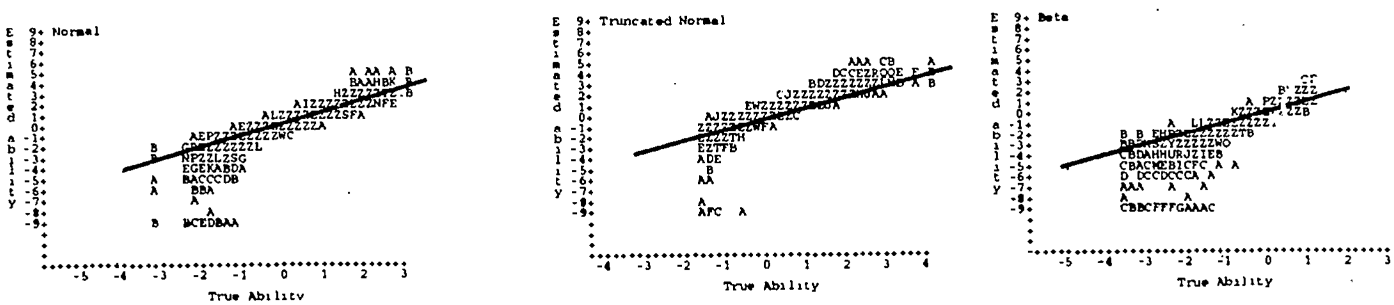

Figure 27. Scatterplots of JML estimates of Θ for 60 items and 1000 examinees.